CEME
Technical Report

CEMETR-2015-05
NOVEMBER-2015

The Center for Educational Measurement and Evaluation

Overview of Factor Analyses Conducted Using Scores from the GOLD Assessment System

Richard Lambert
Do-Hong Kim
Diane Burts

RICHARD LAMBERT
CHUANG WANG
MARK D'AMICO
SERIES EDITORS

A PUBLICATION OF
THE CENTER FOR
EDUCATIONAL
MEASUREMENT
AND EVALUATION

College of Education
UNC CHARLOTTE

CEME
The Center for Educational Measurement and Evaluation

Overview of Factor Analyses Conducted Using Scores from the GOLD Assessment System

Richard Lambert, Do-Hong Kim, and Diane Burts

Center for Educational Measurement and Evaluation

The University of North Carolina at Charlotte

November, 2015

Our research team has conducted numerous factor analyses over many years using data collected using the GOLD assessment system. These analyses have been conducted to both examine and confirm the underlying structure, or dimensionality, of the information provided by the items. These analyses have been conducted using nationally representative norm samples and have included children across the age range for which the measure is intended.

These analyses have included four specific strategies:

1. Exploratory factor analyses (EFA)
2. Confirmatory Factor Analyses (CFA)
3. Measurement invariance studies (MI)
4. Principal Components Analyses of Residuals (PCAR)

EFA is conducted to initially examine how the items tend to cluster together to measure one or more underlying dimensions. CFA is conducted to impose an existing structure on the items to determine if the theoretically derived factor structure as intended by the authors of an assessment can be supported by statistical evidence. MI studies are conducted to examine whether the underlying structure of an instrument varies over time or across sub-groups. PCAR is conducted to examine whether a given set of items measures a single underlying construct.

All of these analyses have demonstrated consistently, with multiple nationally representative samples, that the information provided by the GOLD items can be organized into six scores, each of which corresponds to a specific domain of child development. These six scales are: Social Emotional, Physical, Language, Literacy, Cognitive, and Mathematics development.

The following summary offers the results of a recent study in which we examined whether the six factor structure remains consistent across fall, winter, and spring assessment checkpoints. It also includes an analysis of whether the six factor structure can be maintained across different age groups. The results demonstrate clearly that not is the six factor structure of GOLD confirmed, but remains constant across ages of children and assessment periods.

***Teaching Strategies GOLD®* overview.** *GOLD®* is designed to measure a child's progress in the major developmental and content areas for children ages birth through kindergarten (Heroman et al., 2010). It is intended for use with typically developing children, children with disabilities, children who demonstrate competencies beyond typical developmental expectations, and dual language learners. The assessment tool can be used in early education programs that incorporate the Teaching Strategies curricula as well as in programs which do not use the curricula (Teaching Strategies, LLC, n.d.).

The 38 *GOLD®* objectives and accompanying rating scale items help teachers focus the assessment process as they regularly gather child information through observations, conversations with children and families, samples of children's work, photos, video clips, recordings, etc. The assessment information is to be used in planning appropriate experiences, individualizing instruction, and monitoring and communicating child progress to families and other stakeholders. Data may also be used to help teachers ascertain when additional information or more specific evaluation is needed. As such, *GOLD®* is not a test, nor is it intended to be used as a diagnostic, clinical, or high-stakes instrument.

Although *GOLD®* is similar in some ways to other authentic measures, the tool adds unique contributions to the validated authentic assessment measures currently used. For example, the ability to use a single instrument to assess children from birth to 71 months, rather than having several different measures (e.g., High/Scope Educational Research Foundation, 2002; Meisels et al., 1995; Meisels et al., 2010; Meisels et al., 2008; Schweinhart et al., 1993) has benefits. One instrument which assesses the same broad objectives throughout the early childhood period with appropriate progressions, can be especially beneficial for tracking development and learning longitudinally. It can also assist with program continuity (Snow & Van Hemel, 2008) when children move from one classroom to the next because teachers are already familiar with the assessment system and objectives.

The broader range of item-level rating scale points and behavioral anchors in *GOLD*® (i.e. 10 levels) than those in *COR* (5 levels), *Work Sampling* (3 levels), and *CAR* (2 categories) helps to decrease the likelihood of floor and ceiling effects and can provide useful instructional information for teachers. Indicator and "in between levels" in *GOLD*® allow for additional rating scale points and steps in the progression. The levels in between demonstrate that a child's knowledge, skills, and behaviors are emerging but are not fully established. They help teachers know when to provide support or scaffold child efforts (Early et al., 2010; Vygotsky, 1978). They are also especially helpful for documenting increments of progress for younger children, dual language learners, and children with disabilities. If used as intended, they can assist the teacher in providing appropriate experiences for all children, for planning individualized instruction or specialized small group activities, and knowing when additional information or more specific evaluation is needed (Lopez, Salas, & Flores, 2005).

***GOLD*® development.** Development of *GOLD*® occurred over several years. Its publishers originally proposed to revise the three developmental continua (Teaching Strategies LCC, 2001, 2005, 2006) which were being widely used in early childhood programs (Hyson, 2008). Upon further review of the existing measures and new research, the decision was made to develop a completely new assessment instrument.

Feedback from teachers, administrators, consultants, and Teaching Strategies, LLC professional-development and research personnel was used in the development process. Pilot studies with diverse populations were conducted, and a draft of the measure was sent to leading authorities in their respective fields for content review. Revisions were made based on results of the content validation and pilot studies. Final assessment items were selected on the basis of feedback received during the development process; state early learning standards and the *Head Start Child Development and Early Learning Framework* (U.S. Department of Health & Human Services, Administration of Children & Families, Office of Head Start, 2010); and current research and professional literature including literature that identifies which knowledge, skills, and behaviors are most predictive of school success. This process resulted in a measure having a total of 38 objectives with 23 of them in the areas of social-emotional, physical, language, cognitive, literacy, and mathematics. Although *GOLD*® includes objectives in other areas (i.e., science and technology, social studies, the arts, and English language acquisition), they are not included in the analyses reported in this paper.

Objectives in the social–emotional domain involve understanding, regulating, and expressing emotions; building relationships with others; and interacting appropriately in social situations. Social-emotional competence is critical to children's later academic, social, and psychological outcomes (McCabe & Altamura, 2011). When children's interactions and relationships are positive, they are more likely to have positive short- and long-term outcomes (Commodari, 2013; Peisner-Feinberg et al., 2001; Rubin, Bukowski, & Parker, 1998; Smith & Hart, 2002). Self-regulation is a particularly important construct in the social-emotional domain and is related to academic achievement (McClelland & Cameron, 2011; Ponitz, McClelland, Matthews, & Morrison, 2009; Suchoddetz, Trommsdorff, Heikamp, Wieber, & Gollwitzer, 2009). Both self-regulation and social competence predict children's later reading and math skills (McClelland, Acock, & Morrison, 2006).

The physical domain objectives include gross-motor development (traveling, balancing, and gross-motor manipulative skills) and fine-motor strength and coordination. Physical development affects children's emotional development and their school performance (Rule & Stewart, 2002; Son & Meisels, 2006). It can also affect their social and language development as they interact with peers (Kim, 2005).

The language objectives include understanding and using language to communicate or express thoughts and needs. Language comprehension influences other areas of development such as the closeness of teacher-child relationships (Justice, Cottone, Mashburn, & Rimm-Kaufman, 2008). Language has been found to predict reading skills several years later (Snow, Burns, & Griffin, 1998), and aspects of oral language predict reading comprehension (Roth, Speece, & Cooper, 2002). Children without early experiences that support language development show substantial differences in language understanding and use by age three (Strickland & Shanahan, 2004). When dual language learners acquire English proficiency by the end of kindergarten, they have better cognitive and behavioral outcomes throughout the early school years and beyond than children who become English-proficient after kindergarten (Halle, Hair, Wandner, McNamara, & Chien, 2012).

Objectives in the cognitive domain include approaches to learning (e.g., attention, curiosity, initiative, flexibility, problem solving); memory; classification skills; and the use of symbols to represent objects, events, or persons not present. Symbolic thinking is necessary for language development, problem solving, reading, writing, and mathematical thinking (Deloache, 2004; Younger & Johnson, 2004), and children's symbolic substitution during sociodramatic play is related to their later reading and math skills (Hanline, Milton, & Phelps, 2008). Children's ability to classify is important for learning and remembering (Larkina, Güler, Kleinknecht, & Bauer, 2008), and the more knowledgeable they are about a topic, the more likely they are to categorize at a more mature level (Bjorklund, 2005, Gelman, 1998). The way children approach learning has received increased attention in recent years (Hyson, 2008). Children who have positive approaches to learning are more likely to succeed academically (Howse, Lange, Farran, & Boyles, 2003) and to have more positive interactions with peers (Fantuzzo, Perry, & McDermott, 2004; Hyson, 2008) than children who do not exhibit these characteristics.

The literacy objectives incorporate phonological awareness; alphabet, print, and book knowledge; comprehension; and emergent writing skills. Letter/name writing predicts later literacy, and phonological sensitivity; alphabet knowledge and knowledge of print concepts predict later reading, writing, and spelling success (National Early Literacy Panel [NELP], 2008). Preschool children's development in oral language, phonological awareness, and print knowledge is predictive of later reading abilities (Lonigan, et al., 2011). Children who begin school with less phonological sensitivity, familiarity with the basic purposes and mechanisms of reading, and letter knowledge are especially likely to have difficulty learning to read in the primary grades (NELP, 2008; Snow et al., 1998). Letter knowledge and global phonological sensitivity have been found to be predictors of early reading abilities (Lonigan, Burgess, & Anthony, 2000), while vocabulary and print knowledge are predictive of later numeracy (Purpura, Hume, Sims, & Lonigan, 2011).

The mathematics objectives focus on number concepts and operations, spatial relationships and shapes, measurement and comparison, and pattern knowledge. Children enter school with "everyday" mathematics abilities (Ginsburg, Lee, & Boyd, 2008), and their mathematical skills upon entry to kindergarten are predictive of later reading and math achievement (Duncan et al., 2007). Children's spatial sense is important to other aspects of mathematics, and children with a strong spatial sense tend to do better in mathematics than children without a strong spatial sense (Clements, 2004). Their understandings about counting, number symbols, and number operations are fundamental to their success with more complex mathematics (Ginsburg & Baroody, 2003; Zur & Gelman, 2004).

Previous *GOLD®* research. Several researchers have examined the psychometric properties of *GOLD*® for its use with children representing different ethnic, racial, language, functional status, and age groups. These initial studies, summarized as follows, suggest that *GOLD*® is a psychometrically promising instrument which has utility for children representing diverse populations.

Using a small sample of infants through children two years of age (n = 290), high internal consistency reliability of *GOLD*® (*α*s = .95 - .99) was found (Kim & Smith, 2010). Rasch reliability statistics were moderately high (person separation = 9.42, item separation = 19.20, person reliability = .99, item reliability = .99).

Another study looked at the validity of *GOLD*® for assessing children with disabilities and those for whom English is not their first language. Assessment information was collected on 3-, 4-, and 5-year-old children at the fall (n = 79,324), winter (n = 132,693), and spring (n = 50,558) checkpoints. Differential Item Functioning (DIF) analysis indicated that in general, teachers' ratings were similar for children of similar abilities, regardless of their subgroup membership (Kim, Lambert, & Burts, 2013).

Associations of teacher ratings with child demographics (e.g., age, gender, disability status, English language status) and classroom composition characteristics (e.g., class mean age and percentage ELLs, children with disabilities, and males) were examined with a sample of 21,592 children ages 12 months through 59 months. Using three-level growth curve modeling, findings indicated that teachers' *GOLD*® ratings were associated in anticipated directions for both child and classroom characteristics (Lambert, Kim, & Burts, 2014).

The dimensionality, rating scale effectiveness, hierarchy of item difficulties, and the relationship of *GOLD*® developmental scale scores to child age have also been examined. Data from a norm sample (n = 10,963) of children ages birth to 71 months were analyzed using the Rasch Rating Scale Model. Support was found for the unidimensionality of each domain (i.e., items in each scale measure one and only one underlying latent construct). Results further indicated that teachers can make valid ratings of the developmental progress of children across the measured age range. Correlations were moderately high between each of the scale scores and child age in months, with correlation coefficients ranging from .67 to .73 (Kim, Lambert, & Burts, in press).

Using a different sample of preschool children, researchers examined the relationships between *GOLD*® scale scores and teacher ratings of children's social functioning and learning behaviors and child performance on individually administered direct assessments of academic skills (Lambert, Kim, & Burts, 2013). The sample (n = 299) was diverse and included children attending 51 different Head Start, public pre-k, and private school classrooms across 16 centers in the Northeast United States. In general, the correlations of the external measures with the *GOLD*® domains were moderate and in expected, aligned areas.

Concurrent validity was also examined by researchers in Washington state (Soderberg, Stull, Cummings, Nolen, McCutchen, & Joseph, 2013). Using a modified version of *GOLD*® (i.e., *WaKIDS*) with kindergarten children (n = 333), moderate correlations (r =.50 - .64) with a battery of established norm-referenced achievement instruments were found for the Language, Literacy, and Mathematics areas.

The 23 *GOLD*® objectives included in the studies described in this paper are operationalized into 51 rating scale items: social–emotional (9 items), physical (5 items), language (8 items), cognitive (10 items), literacy (12 items) and mathematics (7 items). Teachers use the information collected on each child to rate their skills, knowledge, and behaviors along a 10- point progression of development and learning from "Not Yet" (Level 0) to Level 9 (exceeds kindergarten expectations). Collected documentation evidence (e.g., observations, artifacts, video recordings, portfolios) is summarized at three checkpoints throughout the year (fall, winter, spring). Levels 2, 4, 6, and 8 on the progressions are "Indicators" and include varied examples from everyday situations that give teachers guidance of what the evidence may look like with majority and with subgroups of children. Levels 1, 3, 5, 7, and 9 are "In-between" levels and do not include examples. They allow for additional steps in the progression as the child demonstrates that skills are emerging in a particular area but are not fully established. Overlapping, color-coded bands indicate the typical age and/or grade-level (i.e., kindergarten) ranges for each item measured.

Prior to the beginning of both the main study and the concurrent validity study, teachers participated in a two-day training on *GOLD®* provided by Teaching Strategies, LLC professional development personnel. Training focused on an overview of the measure and an examination of the objectives and child progressions for development and learning (birth through kindergarten). Teachers watched video clips, examined artifacts, evaluated child portfolios, and participated in large-group discussions related to assessment items. They also completed family conference forms and practiced uploading documentation samples, observational notes, and entering progress checkpoint data online.

Confirmatory factor analysis of cross-sectional data. To address the first research question, the factorial structure of the *GOLD®* was examined using confirmatory factor analysis (CFA) in Mplus (Muthén & Muthén, 1998-2010). The first norm sample data was used for CFA. Given its basis in developmental theory, a six-factor model at the item level that corresponds to the designed structure of the instrument was examined. The chi-square test can be used to evaluate model fit. However, given the sensitivity of this test to sample sizes, alternative goodness-of-fit indices were used to evaluate model fit including Standardized Root Mean Square Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI). RMSEA values near or below .05 are considered a good fit, values between .05 and .08 indicate reasonable fit, and values higher than .10 are unacceptable (Browne & Cudeck, 1993). More recently, RMSEA values close to 0.06 or below and SRMR values close to 0.08 or below were recommended as a good fit (Hu & Bentler, 1999). Generally, CFI values close to or higher than .90 indicate acceptable model fit (Hu & Bentler, 1998). More recently, a more stringent criterion of CFI values close to or higher than .95 has been recommended (Hu & Bentler, 1999).

Results of the CFA showed a statistically significant chi-square statistic ($\chi^2_{1209} = 58213.239, p < .01$), which is not surprising given this study's large sample size. The six-factor model fit the data reasonably well, as evidenced by SRMR = .033, CFI = .932, and RMSEA = .066. The standardized factor loadings for this model are provided in Table 3. All factor loadings were generally large and statistically significant at $p < .001$. The correlations between the six scales were also large (values ranged from .786 to .958) and statistically significant at $p < .001$.

CFA was also examined separately for two age groups (Birth-Age 2 and Ages 3-5). For both age groups, the six-factor model fit the data reasonably well (for children from Birth-Age 2, SRMR = .053, CFI = .903, and RMSEA = .073; for children from Ages 3-5, SRMR = .038, CFI = . 918, and RMSEA = .061). The standardized factor loadings are also provided in Table 3. All factor loadings were generally large (for children from Birth-Age 2, values ranged from .737 to .923; for children from Ages 3-5, values ranged from .739 to .893) and statistically significant at $p < .001$. The correlations between the six scales were also large (for children from Birth-Age 2, values ranged from .702 to .952; for children from Ages 3-5, values ranged from .676 to .932) and statistically significant at $p < .001$.

Longitudinal measurement invariance CFA. The second norm sample longitudinal data were used to address the second research question concerning longitudinal measurement invariance. Prior to conducting longitudinal measurement analyses, CFA was conducted separately on each of the three checkpoints (fall, winter, and spring). The chi-square model fit statistic was statistically significant for all three checkpoints ($\chi^2_{1209} = 107539.136$ for the fall checkpoint, $\chi^2_{1209} = 100823.909$ for the winter checkpoint, and $\chi^2_{1209} = 108293.661$ for the spring checkpoint, $p < .001$). For the fall checkpoint, the model fit the data reasonably well, as evidenced by

RMSEA=.065 and SRMR=.052. However, the CFI value of .87 did not attain the desired cutoff value of .90 proposed by Hu and Bentler (1998).

Fit values for the winter and spring checkpoint indicated acceptable model fit (RMSEA = .063, CFI=.902, SRMR = .041 for the winter checkpoint; RMSEA = .065, CFI = .915, SRMR = .038 for the spring checkpoint). All factor loadings were generally large and statistically significant at $p < .001$. The correlations between the six scales were also large (values ranged from .561 to .958) and statistically significant at $p < .001$. The standardized factor loadings for the model for all three assessment time periods are provided in Table 3. Judging from these results, the data fit the current six-factor model, and thus we preceded to longitudinal measurement invariance CFAs.

We sought to verify whether the scores obtained when teachers use *GOLD*® measure the intended constructs equivalently across time. Following the recommendations of Vandenberg and Lance (2000), we used the single augmented covariance matrix to input data for testing longitudinal measurement invariance. Tests for measurement invariance were conducted by fitting a series of increasingly restrictive models as follows: configural invariance (an unrestricted baseline model), metric invariance (equal factor loadings across time), scalar invariance (equal intercepts across time), and strict invariance (equal residual variances across time). If a more restricted model shows a significant reduction in fit compared to a less restricted model, the more restricted model is not supported, indicating a lack of invariance. The χ^2 difference test is often used for comparing nested models. However, χ^2 tests are known to be extremely sensitive to large sample sizes such as the current study. Thus, this study reported the χ^2 difference test results, but put more emphasis on alternative model fit indices including RMSEA, SRMR, and CFI. Also, following the recommendation of Cheung and Rensvold (2002), we used a change in CFI values of .01 or greater to indicate a significant difference in model fit for testing measurement invariance. The results of the invariance tests are presented in Table 4. Not surprisingly, the χ^2 test and χ^2 difference test ($\Delta\chi2$) results was statistically significant for all models.

Configural invariance. We first examined configural invariance, i.e., whether the pattern of fixed and free factor loadings was similar across time (Model 1). Configural invariance serves as a baseline model for subsequent tests. For each checkpoint, the first item's factor loading was fixed equal to 1.0 and its intercept was fixed equal to 0 for identification; factor loadings, factor variances, covariances, and means were freely estimated across time; and unique variances of observed variables (items) were freely estimated across time. Also, following the recommendation of Vandenberg and Lance (2000), unique covariances between like items across time were estimated. As shown in Table 4 (Model 1), the RMSEA, SRMR, and CFI fit values showed excellent fit of the data, suggesting that the factorial pattern of *GOLD*® was similar across time.

Metric invariance. After establishing configural invariance, metric invariance was examined. The factor loadings of like items were constrained to be equal across time; the factor variance was fixed to 1 at the first checkpoint for identification, but was freely estimated at the second and third checkpoint. The rest of the specifications remained the same as described previously. As shown in Table 4 (Model 2), the CFI difference (ΔCFI) between Model 1 and Model 2 was smaller than the cut-off criterion of .01 and the SRMR and RMSEA fit values were still acceptable, suggesting that factor loadings for the like items were invariant across time.

Scalar invariance. At the next step, the intercepts of items were constrained to be equal across time to evaluate scalar invariance. The factor variance and mean were fixed to 1 and 0, respectively, at the first checkpoint for identification. The rest of the specifications remained the same as described previously. As shown in Table 4 (Model 3), the ΔCFI was negligible. The SRMR and RMSEA values also showed that the model fit was still acceptable for Model 3. The overall results suggest that the intercept for the like items were invariant across time.

Strict invariance. The last step involved constraining residual variances among like items to be equal across time. The rest of the specifications remained the same as described previously. As shown in Table 4 (Model 4), the ΔCFI was less than .01. The SRMR and RMSEA values for Model 4 showed an acceptable model fit. The results indicated that strict invariance was upheld, implying equal residual variances among like items across time.

Overview of Factor Analyses Using Only WAKIDS Data

We have conducted exploratory factor analyses (EFA) and Principal Components Analyses of Residuals using data collected during the fall of the 2013-14 academic year. It is important to note that these analyses are not based on nationally representative norm data set, nor do they include children from the entire age range for which the assessment is intended. Only WAKIDS kindergarten children were included from the fall assessment period. It is also important to note that the WAKIDS program does not use the full set of GOLD items. Rather a custom subset of items has been selected.

The results of the EFA indicated that four underlying dimensions account for 59.15% of the variance in the item responses. The four factors that emerged are: Social Emotional, Physical, Language, and Cognitive. The Social Emotional factor includes the four WAKIDS social emotional items: 1B, 1C, 2C, and 2D. The Physical factor includes the five WAKIDS physical items: 4, 5, 6, 7A, and 7B. The Language factor includes the six WAKIDS language items plus items 18A and 18C. These items are included on the Literacy scale for the full GOLD measure but do include Language related components (interacting during read aloud and story retelling). All of these results reasonably track the results of previous EFA with the exception of the findings regarding items 18A and 18C.

The major difference between the findings using WAKIDS data and previous EFA analyses involves the Cognitive scale. All of the items from the GOLD Cognitive scale clustered together along with the items from the Literacy and Mathematics scales. Therefore, all of the more academically oriented items clustered together, rather than appearing as separate Cognitive, Literacy, and Mathematics scales.

In contrast, the results of Principal Components Analyses of Residuals (PCAR) off some support for the underlying six GOLD scales. For the Cognitive scale, the underlying dimension accounted for 63.5% of the variance in the item responses. Values of at least 50.0% are generally considered acceptable. The first contrast, or largest secondary dimension, accounted for 12.1% of the variance in the residuals. Values under 5.0% are considered acceptable and therefore this somewhat high value suggests the possibility of a second dimension. The fit statistics were also examined as they indicate whether each item fits with the rest of the respective item set. All of these value were within acceptable limits (infit and outfit mean square values .6 – 1.4). The only exception to this finding was for item 13 which yielded values (1.47 and 1.50) slightly above the acceptable range. The item total score correlations ranged from .75 - .81.

For the Language scale, the underlying dimension accounted for 72.6% of the variance in the item responses. The first contrast, or largest secondary dimension, accounted for 7.6% of the variance in the residuals. The fit statistics were within acceptable limits. The only exception to this finding was for item 10B which yielded values (1.53 and 1.49) slightly above the acceptable range. The item total score correlations ranged from .75 - .84.

For the Literacy scale, the underlying dimension accounted for 59.8% of the variance in the item responses. The first contrast, or largest secondary dimension, accounted for 7.4% of the variance in the residuals. The fit statistics were within acceptable limits. The item total score correlations ranged from .59 - .79.

For the Mathematics scale, the underlying dimension accounted for 65.3% of the variance in the item responses. The first contrast, or largest secondary dimension, accounted for 11.2% of the variance in the residuals. This value is a little high and suggests the possibility of a secondary dimension. The fit statistics were within acceptable limits. The item total score correlations ranged from .76 - .83.

For the Physical scale, the underlying dimension accounted for 56.7% of the variance in the item responses. The first contrast, or largest secondary dimension, accounted for 9.9% of the variance in the residuals. This value is a little high and suggests the possibility of a secondary dimension. The fit statistics were within acceptable limits. The item total score correlations ranged from .67 - .70.

For the Social Emotional scale, the underlying dimension accounted for 71.1% of the variance in the item responses. The first contrast, or largest secondary dimension, accounted for 12.6% of the variance in the residuals. This value is a little high and suggests the possibility of a secondary dimension. The fit statistics were within acceptable limits. The item total score correlations ranged from .79 - .84.

In conclusion, there is a wealth of factor analytic evidences that demonstrate and support the underlying dimensionality of the six scales that were designed by the authors of GOLD. This evidence is based on nationally representative norm samples that include all GOLD items, children from all fifty states, and all age groups. The evidence is also generally positive when only kindergarten children from the WAKIDS program are used and only the WAKIDS subset of items is analyzed.